Lateral Divisive Inhibition in Analog Devices for Dynamic Gain Control

G. William Chapman [§], James A. Boyle [§], T. Patrick Xiao, Nad Gilbert, Sapan Agarwal, Frances S. Chance
Sandia National Laboratories
Albuquerque, New Mexico
Email: {gwchapm, jaboyle, txiao, negilbe, sagarwa, fschanc}@sandia.gov

***Abstract*—Standard approaches to computer vision require inputs that are approximately normalized to a given range. For global parameters such as brightness, this can be achieved by a single dynamical global gain. In comparison, the retinal systems of animals possess a number of mechanisms to dynamically control outputs as a function of *local* input properties. We demonstrate a biologically inspired electrical circuit which dynamically changes the gain of each individual input pixel in a nonlinear manner as a function of the surrounding inputs. Through simulation of a minimally complex implementation, we demonstrate that this circuit has a supralinear effect, resulting in an approximate division by the average local intensity. This functionality enables a scene with a high dynamic range to be effectively compressed into a smaller range, while preserving the contrast needed to resolve spatial features at all relevant brightness scales. The circuit is then incorporated into an artificial neural network to solve a modified image classification task where objects have been obstructed by high intensity occlusions. We demonstrate that standard gain approaches, such as min-max scaling or layer normalization, fail while the local normalization circuits are able to adjust gain in a spatially-dependent manner to recover input information and attain high performance.**

***Index Terms*—Neuromorphic Computing, Gain Control, Machine Learning**

I. Introduction

All visual processing systems require some form of dynamic gain adjustment, which must occur before analog-to-digital conversion. In typical approaches there is a simple gain adjustment, such as min-max scaling or clipping, that occurs post-sensor, but before analog-to-digital conversion. This step is essential, as analog-to-digital converters (ADC) and subsequent digital processing steps have a limited bit-width. Failure to adjust the gain of devices can result in clipping, while scaling can preserve global changes in brightness. However, global dynamic gain can result in catastrophic information loss in scenarios where the image brightness varies by orders of magnitude only in subregions of the field of view.

Neurally-inspired approaches offer a possible alternative approach, based on early local lateral interactions [1], [2]. While standard ANN approaches rely solely on additive interactions, the multiplicative interactions proposed by Heeger can be implemented in neuromorphic approaches [1]. Shunting inhibition, for example, can selectively increase the leak-rate of analog neurons or dendritic compartments to allow a supralinear interaction with additive inputs [3], [4]. Such divisive interactions can mimic the scaling approach of dynamic gain control in a local manner by interacting with a local neighborhood [5], and provide a form of adaptive compression before the ADC.

Gain-Control in Biological Systems: Gain control enables multiple, possibly distinct, computational roles in biological systems. Shunting inhibition has been proposed as a mechanism of normalization, capable of adjusting for additive offsets and multiplicative changes in inputs magnitude [6]. Normalization of inputs has been proposed as a canonical cortical column function [7], and normalized activity as assumed in multiple theories of neural coding including predictive coding [8], and spike-rate representations [9]. Beyond normalization, multiplicative interactions can replace entire populations of linearly interacting neurons [10]. Such approaches have been hypothesized to underlie complex computations such as 'gain-fields' to support motor-cortex coordinate transformations [11], [12], complex conjunctive representations for memory-driven behavior [13], or modulate the overall speed of neural processing [14]. More complex multiplicative interactions, such as those that arise with spatially distinct dendritic arborizations, enable additional functionality such as context-dependent processing [15], or attention-driven binding [16], [17]. Finally, while simple models of population encoding assume linearly modulated firing rates with respect to an underlying signal, biological neurons have a maximum sustainable firing rate. This saturating behavior effects population-level which affects population-level encoding and decoding [18]. Multiple learning rules are also thought to rely on 'abnormal' firing rates [19], [20], giving further importance to approximately normalized spike rates during periods of stable stimuli.

Simple Shunting Mechanisms: As an illustrative example, Figure 1 shows the firing rate of simple leaky integrate and fire (LIF) neurons receiving constant current input, and with varying time constants. The units follow the dynamics of Equation 1, with no after-hyperpolarization, refractory period, or other such dynamics.

[§]Equal contribution

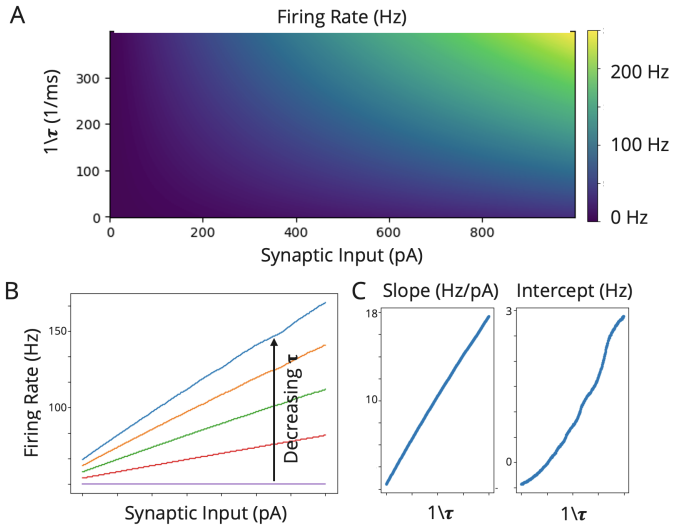

Fig. 1. The effect of tunable time constants in the leaky-integrate and fire dynamics of Equation 1. **A)** For a constant synaptic input current firing rate increases as the time constant (τ) is decreased. **B)** Shows the same results, but as a standard current-frequency curve (FI curve) for various time constants. **C** Linear first to the FI show an approximately linear increase in slope for decreasing τ, indicating a multiplicative effect of synaptic input and τ. The intercept term represents a non-ideal component that downstream mechanisms must accommodate.

$$\tau_v \frac{dv(t)}{dt} = -v(t) + I \qquad S(t) = v(t) \geq 1 \tag{1}$$

Where v is the sub-threshold voltage, I is the input current, τ_v is the time constants of the sub-threshold voltage, and S is the spike. Following a spike, there is a hard reset mechanism:

$$v(t+1) = \begin{cases} 0 & \text{if } S(t) \\ v(t) + dv(t)dt & \text{otherwise} \end{cases} \tag{2}$$

This simple single-compartment, neuromodulator-free model exhibits a firing rate increase that is approximately linear in response to the product of the time constant and input current (Fig 1B). Notably, due to the lack of more complex dynamics the effect of time-constant on firing rate is nearly linear. Additional neuron dynamics would increase the non-ideality of the multiplicative effect, but this may also provide additional computational complexity of individual neurons [21].

In this work: we demonstrate a simple analog circuit which implements a tuneable time constant similar to the example above, but without the spiking behavior. This circuit is a potential candidate for placing in-sensor for localized dynamic gain control. Such a circuit may compress the range of inputs, decreasing ADC resolution, or may be combined with additional circuitry to implement a spiking neural network that replaces the ADC [22]. We demonstrate that this local dynamic gain control can improve the performance of ANNs in the presence of high range inputs.

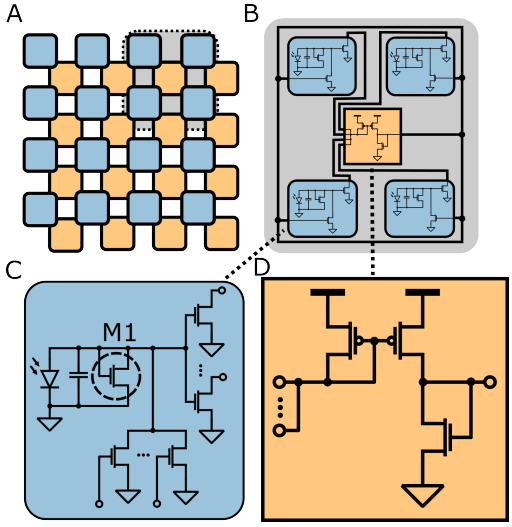

Fig. 2. **Divisive Gain Circuit A)** Input pixels are processed by an analog circuit (grey block) in a convolutional manner. 2×2 is shown for illustration, but 3×3 is used in the simulations. **B)** Each pixel has a current summer (yellow) which sums the activation of neighboring pixels (blue). **C)** Within each pixel, the input is received through a photodiode to charge the capacitor. The capacitor discharges through the transistor M1, which implements self-inhibition. The capacitor also discharges through transistors along the bottom, whose conductances are modulated by inputs from neighboring pixels, and thus implement mutual inhibition among the pixels. The capacitor's state is communicated to other pixels through the transistors on the right, which feed into the current summer. **D)** Pixels are coupled through a current-mirror and summer.

II. Circuit Model

Circuit Design: Our circuit design is presented in Figure 2, and is similar to previous implementations of shunting inhibition, including retinal [23] and dendritic compartments [3]. Individual pixels each contain a photodiode which charges a local capacitor at a rate proportional to the incident light intensity, accumulating a voltage signal. A sub-threshold transistor regulates the rate of charge leakage from the capacitor (Fig 2C). To implement mutual inhibition among pixels, the capacitor charge additionally leaks at a rate that is modulated by the average state of all pixels within a local region (2×2 pixels in Fig. 2). This dynamic leak-rate adjustment is implemented through a current mirror circuit that couples together the neighboring pixels.

Circuit Characteristics: In contrast to the static time constant of the abstract LIF in Figure 1, the dynamic gate voltage in our circuit allows a leak rate which changes over time. The rate of charge accumulation should therefore be a supralinear effect of the input current from the photodiode, and the inverse of neighboring pixel voltages. Figure 3A shows transient responses for three pixels receiving input pulses of a constant magnitude. These transient responses highlight the effect of time-constant, particularly the rate of decay, as neighboring pixels receive non-zero inputs. This effect can be generalized to an increasing number neighboring pixels, as shown in Figure 3, which measures the final voltage as the number of neighboring pixels receiving a concurrent input

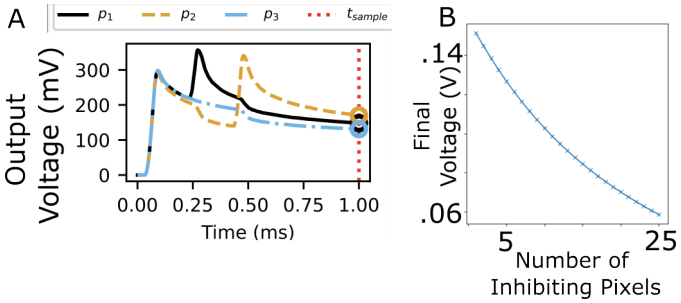

Fig. 3. **Circuit Dynamics A)** Transient response for three pixels (e.g. Fig 2C). All three pixels receive an impulse at $T = 0.1$, while only p_1 receives input at $T = 0.3$, and only p_2 receives input at $T = 0.5$. **B)** Response of a single pixel (as in panel A) as the number of neighboring pixels receiving input is increased. The effect is approximately linear.

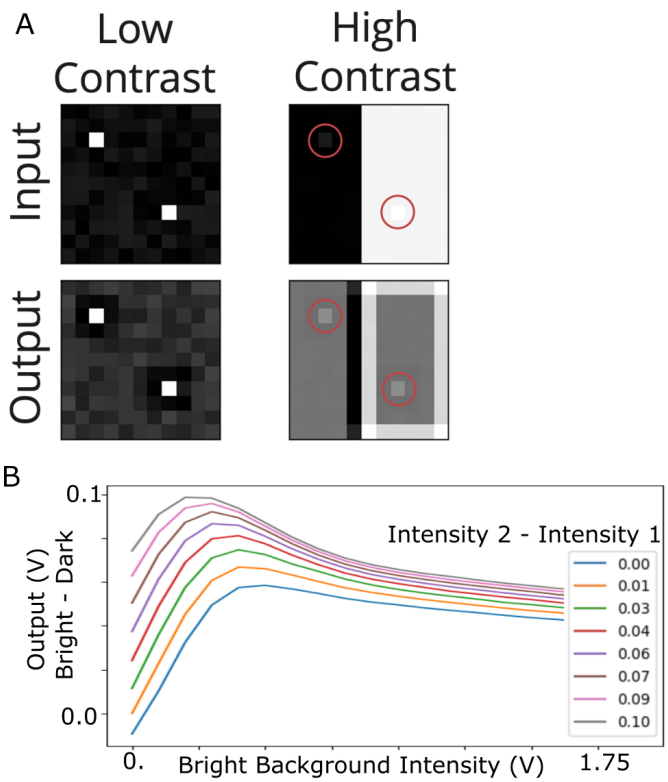

Fig. 4. **A)** *Top* Example of a 2D input to test background-suppression capabilities. For the high-contrast example (top-right), the left hemisphere has a low mean background (0V), while the right hemisphere has a higher (0.5V) mean. Superimposed on both sides is a 0.1V intensity pixel (circled in red). Low-contrast (left) has a zero-mean background on both hemispheres. *Bottom* Corresponding Output of an ideal 3x3 Kernel circuit. Proportional output intensity of the bright spots is decreased compared to the input, with a .05V difference between circled pixels. **B)** Output behavior of the simulated circuit for the same stimulus as above, varying the intensity of the bright background (x-axis) and intensity of bright pixels in the input (lines).

pulse increases. This effect highlights the divisive effects of neighboring inputs for a simple scenario.

For computer vision tasks, we assume that input images are held constant over a period of time (1 ms), and that the pixel voltages at the end of this period represent a 'frame'. A simple input representing the behavior of our circuit under such circumstances can be seen in Figure 4 where, as an example, a single-pixel stimulus is superimposed on a light or dark background, with a 3x3 kernel. Figure 4A shows an example of such inputs under low-contrast, where left and right portions of the image have similar intensity backgrounds, and high contrast, where the left and right portions of the image have a much higher difference in mean magnitude. The bottom row of this panel demonstrates the circuit output, where the difference in voltage of the bright pixels is decreased compared to the input. Figure 4C shows the behavior of this input while sweeping the intensity of the bright-hemisphere background, and intensity of superimposed pixels. An ideal output would have parallel lines, with zero slope as a function of background intensity, and y-values equal to the ground-truth difference in superimposed intensity. Instead, the output is clearly non-linear, with a peak in mid-contrast backgrounds. However, for either low-contrast or high-contrast backgrounds, the output signal decreases the difference in output intensities compared to inputs.

Figure 3B and Figure 4 demonstrate two methods of encoding inputs, either as impulse trains or voltage encoding respectively. Both results show an approximate divisive effect on a given pixel by the number or intensity of surrounding pixels. The degree of nonideality varies depending on the range of inputs and mechanism of encoding, but may be mitigated by limiting the range of operation or training downstream tasks to be aware of these effects. In the next section we assume a voltage-encoding, as in Figure 4, with an ideal divisive interaction, but address these approximations in the Discussion.

III. Machine Learning Experiments

Task: We construct a modified version of the CIFAR-10 task to introduce brightness perturbations, thereby motivating the need for dynamic gain control. Each 8-bit image is first mean-centered at 0.5 and then normalized by dividing by 2.5 times the standard deviation of pixel intensities, computed independently for each channel across the entire dataset. This procedure yields a distribution in which approximately 1.9% of pixel values fall outside the [0,1] range, although all values remain quantized to 256 equally spaced levels. To simulate varying lighting conditions, we apply one of four brightness perturbation conditions, defined along two orthogonal axes: global versus local, and additive ("source") versus multiplicative ("illumination"). Global perturbations are applied uniformly across the entire image, whereas local perturbations are spatially localized using a Gaussian mask centered at a random location, with a standard deviation of 7.5 pixels (approximately one-quarter of the image width). Additive perturbations are sampled from a uniform distribution over [0,1] and added independently to each color channel. These manipulations simulate the presence of an additional light source within the visual scene. Multiplicative perturbations are

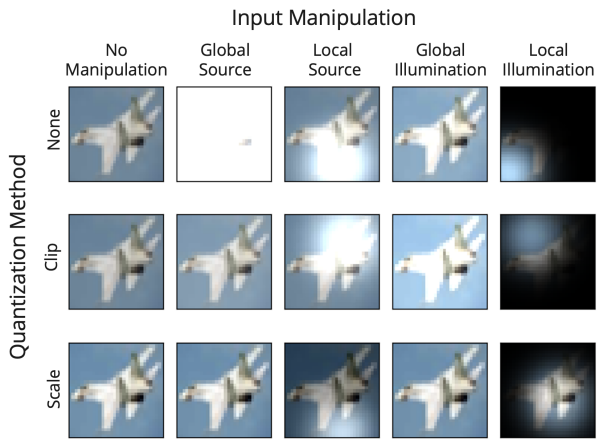

Fig. 5. Example image and manipulations. Each column indicates a modification (none, global addition, local addition, global multiplication, local multiplication), and each row indicates a quantization method (none, hard clipping, or min-max scaling). Note that the first row, despite having no clipping, appears clipped when visualized. For global distractions the clip approach results in washed out features, but the global min-max scaling (bottom) resembles the original image. For local distractions, the clip approach masks the point of interference, while the min-max scaling suppresses the surrounding areas.

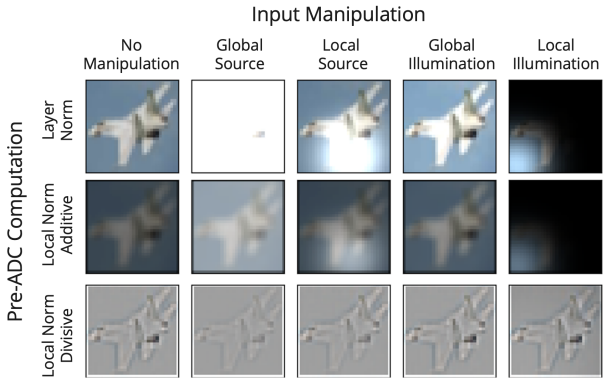

Fig. 6. Example inputs and outputs of pre-ADC normalization circuits, after kernel optimization. (Middle) The additive local normalization approach results in a kernel which subtracts a weighted local mean intensity. The resulting images are overall darker, decrease global or local source perturbations, but the final local illumination images are not recovered. (Bottom) The local multiplicative kernel, contrasted to the local additive approach, does not learn to reconstruct the unperturbed images. Instead, outputs appear to extract and enhance boundaries, resulting in a recognizable outline in all five conditions.

sampled uniformly from [1,2] and applied identically across all channels, simulating a global increase in luminance. Both forms of perturbation expand the dynamic range of the input by a factor of two, underscoring the challenge posed to static normalization schemes.

Local Normalization Layers: Perturbed inputs then go through one of three a normalization operations. The first approach is layer normalization, which is used widely in machine learning [24]. This approach normalizes each channel's input by subtracting the mean and dividing by the standard deviation on a per-sample basis. Layer normalization can not be implemented in the circuit of Figure 2, but represents a baseline comparison, and could be implemented with a more complex approach. The second two variants represent analog in-sensor gain control, which could be implemented as variations of our circuit. The second approach ('LocalNorm-A') applies a local, linear, radially symmetric convolution [25], which is a trainable operation similar to standard on-center off-surround receptive fields. The final approach ('LocalNorm-D') implements an idealized approximation of the circuit dynamics described in section II, by applying a radially-symmetric *divisive* convolution:

$$Y = \frac{X}{X * K} \tag{3}$$

where $*$ is the convolution operator and K is the radially symmetric kernel. For the purposes of this work we consider a 3x3 kernel with stride of 1.

Quantization: The normalized images are subsequently subjected to one of three simulated analog-to-digital conversion (ADC) operations. In the 'none' condition, channel values are zero-centered but retain their full floating-point precision, modeling an idealized ADC with no quantization. This condition serves as a baseline for evaluating a network's ability to process inputs with increased dynamic range or localized occlusions, independent of quantization-related information loss. The 'clip' condition simulates a static ADC with 256 quantization bins uniformly distributed over the fixed range $[0, 1]$. As a result, any pixel values falling outside this range due to brightness perturbations are clipped, leading to a complete loss of information in those regions. In contrast, the 'scale' condition implements a form of global dynamic gain control by linearly rescaling pixel values such that the minimum and maximum intensities within each image map to 0 and 1, respectively. Quantization bins are then spaced uniformly across this adjusted range. While this approach prevents clipping, it increases the width of each bin and thereby reduces the effective resolution of the quantized representation.

Network & Training: Following analog pre-processing and quantization, the resulting pixel values were run through an artificial neural network with the VGG-11 architecture [26], ReLU activation functions, and cross-entropy categorization loss. Except where noted, all network initializations and hyperparameters followed the PyTorch defaults. Training used stochastic gradient descent (momentum = 0.9, weight decay = 5×10^{-4}) with a cosine annealing schedule for the learning rate, which was initialized at 0.1 and decreased over 200 epochs. All reported results are the average of 10 independent initializations and training with the given set of parameters. Quantization layers were implemented with surrogate gradient descent [27]. During the forward pass the quantized values were used, and during the backward pass the straight-through estimator (STE) was used to estimate local gradients with respect to unquantized values.

IV. Results

The performance of all network configurations, averaged across runs, is summarized in Table IV.

Normalization	Quantization	Dataset perturbation				
		Base	Global Source	Global Illumination	Local Source	Local Illumination
None	None	91.3	90.3	90.2	90.8	76.7
	Clip	91.4	12.4	10.7	79.8	76.5
	Scale	91.1	19.9	91.3	80.8	80.2
LayerNorm	None	91.5	91.3	91.2	13.7	10.0
	Clip	91.4	88.7	88.6	13.6	11.1
	Scale	91.5	89.9	91.7	15.2	72.6
LocalNorm-A	None	91.1	90.5	91.2	90.3	10.0
	Clip	91.2	91.5	91.4	91.2	69.2
	Scale	90.9	91.0	91.0	91.3	75.2
LocalNorm-D	None	90.7	90.9	90.2	90.1	89.2
	Clip	90.8	90.7	90.9	90.9	91.2
	Scale	91.6	90.5	91.2	90.8	88.2

TABLE I
PERFORMANCE ACROSS NORMALIZATION-QUANTIZATION CONFIGURATIONS FOR EACH MANIPULATION. CELLS IN RED PERFORM AT LESS THAN 50% ACCURACY, AND THOSE IN YELLOW PERFORM BETWEEN 50% AND 85%.

Global Manipulations: The baseline classification accuracy for this network is 91.3%, roughly in line with non-augmented approaches in the literature for CIFAR-10, and all network variations achieve 90-92% accuracy on the unmodified CIFAR-10 task ('Base' dataset in the table). Floating-point networks (quantization 'None') are able to classify images with global additive and multiplicative perturbations with >90% accuracy, indicating that the internal batch normalization layers of VGG networks are sufficient to accommodate the increased dynamic range of the inputs. When analog-to-digital quantization is applied, without any normalization pre-processing, the networks fail to converge and perform at nearly the accuracy of random chance. Analog layer normalization and both forms of local normalization are able to overcome this effect, however. This indicates that the dataset perturbations result in quantization errors that affect network performance, but which can be overcome with appropriate pre-quantization processing.

Local Manipulations: While most network configurations perform well in the global manipulation tasks, the classification accuracy is more varied when local manipulations are present. Local additions disrupt quantized no-normalization and layer-normalization networks. This can be understood intuitively, as local additions will occlude portions of images with 'clip' quantization, and decrease the resolution in images with globally normalized ('scale') quantization. Both additive and multiplicative local normalizations were able to correct for local additive effects. Only the local multiplicative normalization which implements Equation 3, or equivalently an idealized version of the circuit in Figure 2, was able to correct for local multiplicative alterations.

V. DISCUSSION

This work proposed a compact circuit implementation of local divisive gain, along with characteristic curves for its spatial and temporal response. We demonstrated that an ANN-based algorithm is able to utilize idealized circuit behavior to implement intelligent pre-ADC operations to overcome high range of inputs in a modified classification task.

Applications: Our object recognition task utilized four manipulations which compared local versus global changes and multiplicative versus additive brightness. The global multiplicative case is most directly relevant to biological systems, where animals are able to navigate in environments which have orders of magnitude absolute range in brightness. However, the local approaches, both additive and multiplicative, are more relevant for artificial approaches. For instance, a large field of view may contain local light sources that do not equally illuminate the entire field (local multiplicative). In this work, we considered the case of images which change at a rate substantially lower than the intrinsic dynamics of the gain circuit. However, biological systems evolve on a timescale closer to external stimuli, and have been shown to increase sensitivity to spatiotemporal patterns of interest in a supra-linear manner, while suppressing background patterns [28]. Such approaches could be incorporated to future applications to increase sensitivity to behaviorally-relevant stimuli, similar to how local background suppression is utilized by our ANN [29].

Hardware Ideality: Notably, we utilized the approximately divisive results of *final* pixel voltage for inputs which are held for extended periods of time. This results in two notable idealizations of the circuit in the machine learning experiment: 1) We implement local normalization as an exact division with trainable shunting, and 2) We ignore any transient activity in the analog circuit, assuming that the external stimuli change relatively slowly compared to the processing speed of the network. Future work may investigate a SPICE-in-the-loop version of the work presented here, through the use of physics-aware training [30]. Such an approach would allow us to incorporate the temporal response of the analog front-end with neural networks that evolve through time (e.g. spiking neural networks) rather than training on idealized analog responses. Furthermore, direct optimization of individual circuit parameters would enable more optimal implementations of divisive gain that can handle an even larger input dynamic range. These include, for example, tunable capacitances to enable trainable time constants [31], optimal

coupling strengths, and others.

ACKNOWLEDGMENT

This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC, a wholly owned subsidiary of Honeywell International Inc. for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. The authors own all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan https://www.energy.gov/downloads/doe-public-access-plan. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

REFERENCES

[1] D. J. Heeger, "Normalization of cell responses in cat striate cortex," *Visual Neuroscience*, vol. 9, no. 2, pp. 181–197, Aug. 1992.

[2] D. J. Heeger and K. O. Zemlianova, "A Recurrent Circuit Implements Normalization, Simulating the Dynamics of V1 Activity," *Proceedings of the National Academy of Sciences*, vol. 117, no. 36, pp. 22 494–22 505, Sep. 2020.

[3] F. S. Chance and S. G. Cardwell, "Shunting Inhibition as a Neural-Inspired Mechanism for Multiplication in Neuromorphic Architectures," in *Proceedings of the 2023 Annual Neuro-Inspired Computational Elements Conference*, 2023, pp. 41–46.

[4] J. Edwards, L. Parker, S. G. Cardwell, F. S. Chance, and S. Koziol, "Neural-Inspired Dendritic Multiplication Using a Reconfigurable Analog Integrated Circuit," in *2024 IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2024, pp. 1–5.

[5] C. A. Mead and M. Mahowald, "A Silicon Model of Early Visual Processing," *Neural Networks*, vol. 1, no. 1, pp. 91–97, Jan. 1988.

[6] M. Carandini and D. J. Heeger, "Summation and Division by Neurons in Primate Visual Cortex," *Science*, vol. 264, no. 5163, pp. 1333–1336, May 1994.

[7] ——, "Normalization as a canonical neural computation," *Nature Reviews Neuroscience*, vol. 13, no. 1, pp. 51–62, Jan. 2012.

[8] R. P. N. Rao and D. H. Ballard, "Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-Field Effects," *Nature Neuroscience*, vol. 2, no. 1, pp. 79–87, Jan. 1999.

[9] C. Eliasmith and C. H. Anderson, *Neural Engineering: Computation Representation and Dynamics in Neurobiological Systems*. Cambridge, MA: MIT Press, 2004.

[10] C. Weber and S. Wermter, "A Self-Organizing Map of Sigma-Pi Units," *Neurocomputing*, vol. 70, no. 13-15, pp. 2552–2560, Aug. 2007.

[11] A. Pouget and T. J. Sejnowski, "Spatial Transformations in the Parietal Cortex Using Basis Functions," *Journal of Cognitive Neuroscience*, vol. 9, no. 2, pp. 222–237, Mar. 1997.

[12] Y. E. Cohen and R. A. Andersen, "A Common Reference Frame for Movement Plans in the Posterior Parietal Cortex," *Nature Reviews Neuroscience*, vol. 3, no. 7, pp. 553–562, Jul. 2002.

[13] A. S. Alexander, J. C. Tung, G. W. Chapman, A. M. Conner, L. E. Shelley, M. E. Hasselmo, and D. A. Nitz, "Adaptive Integration of Self-Motion and Goals in Posterior Parietal Cortex," *Cell reports*, vol. 38, no. 10, p. 110504, 2022.

[14] D. Wyrick and L. Mazzucato, "State-Dependent Control of Cortical Processing Speed via Gain Modulation," *arXiv:2004.04190 [q-bio]*, Apr. 2020.

[15] M. Baronig and R. Legenstein, "Context Association in Pyramidal Neurons through Local Synaptic Plasticity in Apical Dendrites," *Frontiers in Neuroscience*, vol. 17, p. 1276706, 2024.

[16] A. Treisman, "Feature Binding, Attention and Object Perception," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 353, no. 1373, pp. 1295–1306, Aug. 1998.

[17] M. Lundqvist, J. Rose, P. Herman, S. L. L. Brincat, T. J. J. Buschman, and E. K. K. Miller, "Gamma and Beta Bursts Underlie Working Memory," *Neuron*, vol. 90, no. 1, pp. 152–164, 2016.

[18] J. R. Hinman, M. P. Brandon, J. R. Climer, G. W. Chapman, and M. E. Hasselmo, "Multiple Running Speed Signals in Medial Entorhinal Cortex," *Neuron*, vol. 91, no. 3, pp. 666–679, 2016.

[19] C. Clopath and W. Gerstner, "Voltage and Spike Timing Interact in STDP - a Unified Model," *Frontiers in Synaptic Neuroscience*, vol. 2, no. JUL, pp. 1–11, 2010.

[20] M. E. Rule and T. O'Leary, "Self-Healing Codes: How Stable Neural Populations Can Track Continually Reconfiguring Neural Representations," *Proceedings of the National Academy of Sciences*, vol. 119, no. 7, p. e2106692119, Feb. 2022.

[21] R. Naud, N. Marcille, C. Clopath, and W. Gerstner, "Firing patterns in the adaptive exponential integrate-and-fire model," *Biological Cybernetics*, vol. 99, no. 4-5, pp. 335–347, Nov. 2008.

[22] M. P. E. Apolinario, A. K. Kosta, U. Saxena, and K. Roy, "Hardware/Software Co-Design With ADC-Less In-Memory Computing Hardware for Spiking Neural Networks," *IEEE Transactions on Emerging Topics in Computing*, vol. 12, no. 1, pp. 35–47, Jan. 2024.

[23] C. Mead, "Neuromorphic Electronic Systems," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1629–1636, Oct. 1990.

[24] J. L. Ba, J. R. Kiros, and G. E. Hinton, "Layer Normalization," Jul. 2016.

[25] W. Fuhl and E. Kasneci, "Rotated Ring, Radial and Depth Wise Separable Radial Convolutions," Jan. 2021.

[26] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[27] I. Hubara, M. Courbariaux, D. Soudry, R. El-Yaniv, and Y. Bengio, "Quantized Neural Networks: Training Neural Networks with Low Precision Weights and Activations," *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6869–6898, 2017.

[28] N. Pirogova and A. Borst, "Contrast normalization affects response time-course of visual interneurons," *PloS one*, vol. 18, no. 6, p. e0285686, 2023.

[29] M. Yedutenko, M. H. C. Howlett, and M. Kamermans, "High Contrast Allows the Retina to Compute More Than Just Contrast," *Frontiers in Cellular Neuroscience*, vol. 14, p. 595193, Jan. 2021.

[30] L. G. Wright, T. Onodera, M. M. Stein, T. Wang, D. T. Schachter, Z. Hu, and P. L. McMahon, "Deep Physical Neural Networks Trained with Backpropagation," *Nature*, vol. 601, no. 7894, pp. 549–555, 2022.

[31] Y. Li, R. Kim, and T. J. Sejnowski, "Learning the Synaptic and Intrinsic Membrane Dynamics Underlying Working Memory in Spiking Neural Network Models," Neuroscience, Preprint, Jun. 2020.